



# **Agilent GeneSpring Software**

## **Data Update Guide**

GeneSpring uses different types of data and information to analyze your sample data in the right framework and put your results in a biological context. This guide provides detailed information on the application of these different types of data in GeneSpring and how you can download, import, and update them.

The following types of data are covered in this guide:

- Technologies
- Annotations
- Gene sets
- GeneSpring Interaction Databases
- BridgeDb and ID Mapper databases



**Agilent Technologies**

## Overview of data types in GeneSpring

To provide the right biological context and framework for your sample data and results, GeneSpring provides the option to import, download, and update many different types of data. The following sections describes these different data types and their application in GeneSpring in detail.

### Technologies

In GeneSpring, technology refers to the microarray type that was used to generate the experiment data. For example, Agilent Whole Genome and Affymetrix HG-U133 PLUS2 are two different array types. Technologies contain biological information about the genes on a specific array type.

Agilent provides technologies for microarrays from Agilent, Affymetrix, Illumina, as well as custom arrays. Create the appropriate technology for each new array type before you start analyzing data from that array. When updates to any of the existing technologies become available, you are notified the next time you start GeneSpring.

You can also augment your technology with the annotations contained in the biological genomes from the National Centre for Biotechnology Information (NCBI).

#### See also

[“How to create a technology”](#) on page 7

[“How to update technology annotations”](#) on page 17

[“How to create a biological genome in GeneSpring”](#) on page 27

## Annotations

In GeneSpring, you can access genome annotations for many organisms and from many different biological databases (Ensembl, RefSeq, etc.). Such annotations, like reference genome builds, transcripts, dbSNP, GC percent, Conservation, HomoloGene Groups, Sure Select targets, etc. are available to download from the Agilent server for the most common organisms.

The Annotations Manager in GeneSpring allows you to manage these annotations and informs you when updates to previously downloaded annotations become available. If your organism is not available to download from the Agilent server, the Annotations Manager provides the facility to create a custom organism and add custom annotations.

Reference and target annotations are specific to whole genome and SureSelect sequence-based experiments. Genome builds, cytoband, and gene and transcript models are used both in sequence-based analyses as well as expression, exon splicing, copy number, and association experiments for visualizing results in the GeneSpring Genome Browser.

### See also

[“How to download annotations in the Annotations Manager” on page 32](#)

[“How to update annotations in the Annotations Manager” on page 36](#)

## Gene Sets

A gene set is a group of genes that share a common function, chromosomal location, or regulation. Gene sets are available to download from the Broad Institute (<http://www.broadinstitute.org/gsea/msigdb/index.jsp>), and Stanford University (<http://www-stat.stanford.edu/~tibs/GSA/>).

Import these gene sets into GeneSpring to perform Gene Set Enrichment Analysis (GSEA) or Gene Set Analysis (GSA).

### See also

[“How to download Broad GSEA gene sets”](#) on page 37

[“How to download gene sets from the Stanford University website”](#) on page 39

[“How to import gene sets into GeneSpring”](#) on page 40

## GeneSpring Interaction Databases

GeneSpring interaction databases contain organism-specific information about proteins and organism-independent information about metabolites at the entity-level as well as the interactions between them. This information is derived from an NLP-based search of PubMed abstracts, interactions from IntAct (<http://www.ebi.ac.uk/intact/>), and the Saccharomyces Genome Database (SGD, <http://yeastmine.yeastgenome.org/yeastmine/aspect.do?name=Interactions>) for yeast.

GeneSpring requires interaction databases for the experiment organisms during Pathway Analysis with BioPAX (Level 2) pathways, pathways created with Pathway Architect or from an imported \*.txt file, and NLP or MeSH derived interaction networks (see ‘Biological Pathways’ on page X).

Interaction databases for Homo sapiens and metabolites are packaged with your GeneSpring installation. In addition, Agilent provides interaction databases for other commonly cited organisms. These databases can be downloaded to GeneSpring through the Agilent server or imported from a GeneSpring update file. GeneSpring Interaction databases are updated on a regular basis and GeneSpring notifies you when an update becomes available.

### See also

[“How to download GeneSpring interaction databases” on page 41](#)

[“How to update GeneSpring interaction databases” on page 45](#)

## BridgeDb and ID Mapper databases

The BridgeDb mapping framework (<http://www.bridgedb.org>) is integrated in GeneSpring to allow you to match entities across annotation sources, for example Entrez Gene and Ensembl. Of the data sources available in BridgeDb, GeneSpring currently supports a restricted number of data sources for metabolites, genes, and proteins for a number of organisms. In addition, custom ID Mapper databases are also available for metabolites and some organisms.

BridgeDb and the custom ID Mapper databases are required in GeneSpring to perform Single Experiment and Multi-Omic Pathway Analysis. You can download these databases for the organisms of your interest directly within GeneSpring from the Agilent server. BridgeDb and ID Mapper databases are updated frequently and GeneSpring notifies you when updates become available.

### See also

[“How to download BridgeDb and ID Mapper databases”](#) on page 46

[“How to update BridgeDb and ID Mapper databases”](#) on page 49

# Creating and Updating Technologies

This section describes the required steps for creating the appropriate technology for your experiment and updating it when updates are available. This section also provides details on how you can augment a technology by adding the information contained in an NCBI biological genome to your technology.

## How to create a technology

There are many ways in which to create a technology in GeneSpring:

- From Agilent Server
- From Agilent eArray
- Custom Agilent miRNA
- From GeneSpring Update File
- From .gpr file
- Custom from file
- Affymetrix Expression

The steps required for each of these options are described in the following sections.

### To create a technology from the Agilent server

- 1 Make sure you are connected to the internet.
- 2 Select **Annotations > Create Technology > From Agilent Server** from the main menu.

The **Automatic Software Update** window opens.

- 3 Use the **Find** function or scroll bar to identify the required technology from the list of available technologies.
- 4 Select the technology. You can also select multiple technologies to create at the same time.

A summary of the selected technologies appears in the bottom pane of the window.

- 5 Click **Update**.

An **Automatic Software Update** information dialog informs you of the amount of hard disk space that is required to create the selected technologies.

- 6 Click **OK**.

GeneSpring downloads the technology from the Agilent Server. An **Automatic Software Update** information dialog confirms when the download is complete.

- 7 Click **OK**.

After GeneSpring configures the new technology, as indicated in the progress bar, technology creation is complete.

You can now proceed with creating your new experiment in GeneSpring.



**To create a technology from Agilent eArray**

To create a technology from the Agilent eArray portal, you require an account with the portal. Use this option to create a technology for custom arrays that you created using the eArray portal.

**NOTE**

Technologies for standard arrays can be created by downloading the necessary details from the Agilent Server as described in [“To create a technology from the Agilent server”](#) on page 8.

- 1 Make sure you are connected to the internet.
- 2 Select **Annotations > Create Technology > From Agilent eArray** from the main menu.  
The **Create eArray Technology** dialog box opens.
- 3 Provide your Agilent eArray **User Name** and **Password** as well as the **AMADID NUMBER** of the microarray for which you want to create the technology.
- 4 Select the experiment organism from the **Organism** drop-down list.
- 5 Select the technology type from the **Technology Type** drop-down list.
- 6 Click **OK**.

GeneSpring downloads the required information from the Agilent eArray portal. Once the download is complete, a **Technology Inspector** opens. you can

- 7 Review the details for the technology, for example the number of entities it contains, the vendor, and organism. Click **OK**.

The technology has been created and you can proceed with creating your new experiment in GeneSpring.

### To create a technology from Custom Agilent miRNA

Use this option to create miRNA technologies for custom Agilent miRNA arrays for which you have a GEML file.

#### NOTE

If you have an Agilent eArray account and created the custom miRNA in the eArray portal, you can also create the technology using the steps described in [“To create a technology from Agilent eArray”](#) on page 9.

---

- 1 Select **Annotations > Create Technology > From Custom Agilent miRNA** from the main menu.

The **Create miRNA Technology from GEML file** dialog box opens.

- 2 Click **Browse** and locate the GEML file on your computer. Click **Open**.
- 3 Select the experiment organism from the **Organism** drop-down list.
- 4 Click **OK**.

GeneSpring extracts all the required information to create the technology from the GEML file. When the process is complete, a **Technology Inspector** opens.

- 5 Review the details for the technology, for example the number of entities it contains, the vendor, and organism. Click **OK**.

The technology has been created and you can proceed with creating your new experiment in GeneSpring.

### To create a technology from GeneSpring Update File

If you were not able to create a technology using one of the other options described in this section, you can contact [GeneSpring technical support](#) for assistance. In most cases they will be able to send you a GeneSpring update file (\*.update, \*.zip) for creating the technology.

- 1 Select **Annotations > Create Technology > From GeneSpring Update file** from the main menu.

The **Please Select A .update File** dialog box opens.

- 2 Locate the \*.update file or \*.zip archive on your computer.
- 3 Click **Open**.

All the technologies in the update file appear listed in the **Automatic Software Update** window.

- 4 Select the technologies you want to create.

A summary of the selected technologies appears in the bottom pane of the window.

- 5 Click **Update**.

An **Automatic Software Update** information dialog informs you of the amount of hard disk space that is required to create the selected technologies.

- 6 Click **OK**.

GeneSpring extracts the technology details from the update file. An **Automatic Software Update** information dialog confirms when the process is complete.

- 7 Click **OK**.

After GeneSpring configures the new technology, as indicated in the progress bar, technology creation is complete.

You can now proceed with creating your new experiment in GeneSpring.

### To create a technology from \*.gpr file

Use this option to create technologies for custom two color microarray experiments for which you have a GenePix Results File (\*.gpr) file.

- 1 Select **Annotations > Create Technology > From .gpr file** from the main menu.

The **Create GenePix Technology** dialog box opens.

- 2 Provide the **Technology name**.
- 3 Select the experiment organism from the **Organism** drop-down list.
- 4 Click **Browse** and locate the \*.gpr file on your computer. Click **Open**.
- 5 Click **Finish**.

GeneSpring extracts the technology details from the \*.gpr file. An **Information** dialog confirms when the process is complete.

- 6 Click **OK**.

The custom technology has been created and you can now proceed with creating your new experiment in GeneSpring.

### To create a custom technology from file

Use this option to create technologies from a tab-delimited file (\*.csv, \*.tsv, \*.xls, \*.txt) for custom single color or two color microarray experiments from vendors for which you cannot download technologies from the Agilent Server.

- 1 Select **Annotations > Create Technology > Custom from file** from the main menu.

The **Create Custom Technology** wizard-driven workflow opens.

- 2 In Step 1 of 9 of the workflow, specify the following details:
  - a Select the **Technology type**. You can only create single color or two color experiment technologies with this option.
  - b Provide the **Technology name**.
  - c Select the experiment organism from the **Organism** drop-down list.
  - d Click **Browse** and locate the technology file on your computer. Click **Open**.
  - e Specify whether there is only one sample or several samples in a single file from **Number of samples in single data file** drop-down list.
  - f **Optional:** Click **Browse** to choose a file that contains additional annotations to be added to the custom technology.

#### NOTE

Additional annotations are not required to create the technology. Without additional annotations, technology creation is complete after Step 4 of 9 of this workflow. The remaining steps are dedicated to correctly importing the information contained in the annotations file.

- g Click **Next >>** to proceed to the next step.
- 3 In Step 2 of 9 of the workflow, provide details about the file format.

The **Preview** table indicates how GeneSpring read the selected input file by default. If the data columns were not recognized correctly, change the settings for the **Separator**, **Text qualifier**, **Missing value indicator**, and **Comment indicator** to match the specifications of the input file. Click **Next >>**.

- 4 In Step 3 of 9 of the workflow, select which rows you want GeneSpring to use for creating this technology:
  - a Under **Row Options**, select whether you want to use all rows, a fixed number of rows between two indices (indicated in the first column of the **Preview** table), or between two marks based on the associated values displayed in the **Preview** table.
  - b Under **Header Row Options**, indicate whether the first row contains headers or not.
  - c Click **Next >>** to proceed to the next step.
- 5 In Step 4 of 9 choose the identifier and signal column:
  - a Choose the column that contains the identifier for the entities from the **Identifier** drop-down list .
  - b Choose the column that contains the background corrected signal values from the **BG Corrected Signal** drop-down list.
  - c **Optional:** Determine the column that contains flag values. Click **Configure** to map the flag values in your file to the flag values used in GeneSpring in the **Configure the Column** dialog box. For Agilent technologies, these flag values are **Detected**, **Not Detected**, and **Compromised**. For non-Agilent technologies these flag values are **Present**, **Marginal**, and **Absent**. Click **OK**.
  - d Click **Finish** to complete technology creation.

If you specified an annotation file in Step 1 of 9 of this workflow, click **Next >>** to continue. Provide the required details to correctly import the annotations and add them to the new technology in the remaining steps.

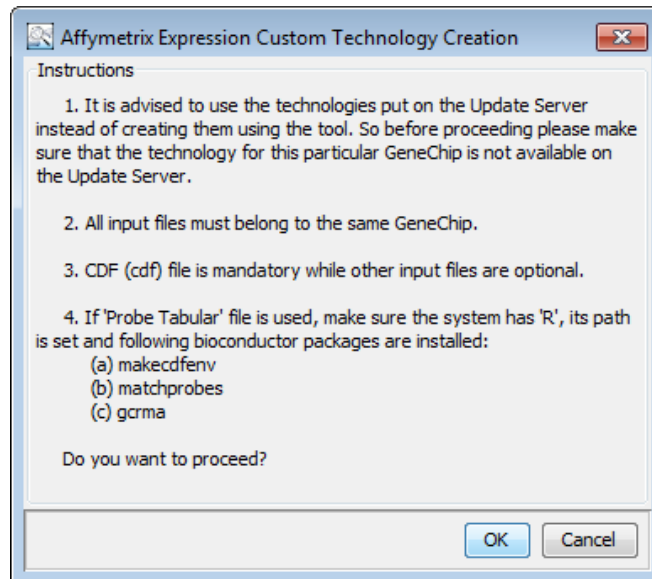
GeneSpring creates the technology, as indicated by the progress bar. You can now create your new experiment.

### To create a custom Affymetrix Expression technology

Use this option to create the technology for a custom Affymetrix Expression microarray from a \*.cdf file. This procedure also allows you to provide additional information through other library files supplied by Affymetrix (\*.csv, \*.psi, \*.cif, Probe Tabular).

- 1 Select **Annotations > Create Technology > Affymetrix Expression** from the main menu.

The **Affymetrix Expression Custom Technology Creation** dialog box opens (see [Figure 1](#)).



**Figure 1** Affymetrix Expression Custom Technology Creation dialog box

- 2 Confirm that your files and GeneSpring installation fulfill all the conditions listed in this dialog box. Click **OK**.
- 3 Choose the input files.

The CDF file is the minimum requirement to create a custom Affymetrix Expression technology. Click **Browse** to locate the \*.cdf file on your computer. Click **Open**.

## Creating and Updating Technologies

If you want to add any of the other optional library files at this time, click **Browse** next to each of the available file types and locate each of them on your computer.

### 4 Click **OK**.

GeneSpring creates the technology from the files you indicated. A **Progress** bar informs you of the progression of technology creation. An Information message confirms when the technology has been created. Click **OK**.

The custom Affymetrix Expression technology is now present in GeneSpring and you can proceed with creating the new experiment.



## How to update technology annotations

The data libraries that hold the information required by GeneSpring to create a technology for standard microarrays change as new information becomes available and is incorporated in the respective libraries. Whenever vendors update the data libraries for their microarray chips, Agilent collects this information for you. Agilent then packages it in such a way that it can be imported into GeneSpring to create a technology (see [“How to create a technology”](#) on page 7) or existing technologies can be updated from the Agilent server as described in [“To update technology annotations from the Agilent server”](#).

You can also update these standard microarray technologies using one of the following options:

- from the Agilent eArray portal (for Agilent standard technologies).
- from a GeneSpring update file, provided by [GeneSpring technical support](#) on request.

To augment an existing custom technology with additional annotations, GeneSpring provides the following options to update technology annotations:

- tab-delimited file
- biological genome from the NCBI

### NOTE

Updating an existing technology does not affect existing experiments and the results (entity lists) it contains. GeneSpring only uses the updated technology for creating new experiments and for any further analyses you perform in an existing experiment.

The steps for each of these options to update or augment an existing technology are described in detail in the following sections.

### To update technology annotations from the Agilent server

Every time you launch GeneSpring and your computer is connected to the internet, GeneSpring connects to the Agilent server to look for updates to existing technologies that you have previously created.

If an update is available, the Automatic Software Update window opens automatically during start up.

#### NOTE

If you do not want to update technologies at start up, you can close the Automatic Software Update window and return to it later by selecting **Annotations > Update Technology Annotations > From Agilent Server** from the main menu.

---

All the updates to the technologies that were previously created are already selected. The details for each technology are displayed in the lower pane of the window.

- 1 Review the list of technologies that was preselected.
- 2 Unselect any technologies you do not want to update.
- 3 Click **Update**.

An **Automatic Software Update** information dialog informs you of the amount of hard disk space that is required to create the selected technologies.

- 4 Click **OK**.

GeneSpring downloads the update from the Agilent Server. An **Automatic Software Update** information dialog confirms when the download is complete.

- 5 Click **OK**.

Your technologies are now updated.

#### NOTE

You can inspect the information associated with the technology in a particular experiment in the Technology Inspector. To open the inspector, right-click the experiment name in the Project Navigator and select Inspect Technology.

---

### To update technology annotations from Agilent eArray

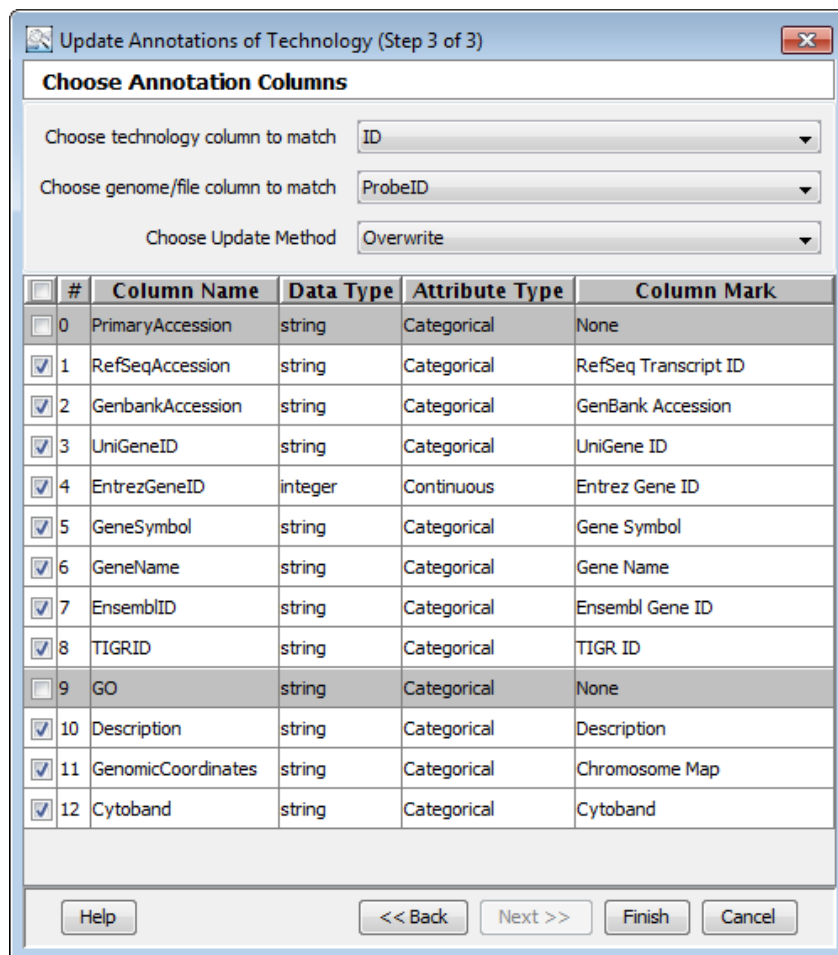
To update technology annotations from the Agilent eArray portal, you require an account with the portal. During this process you are required to provide your aArray account user name and password.

- 1 Make sure you are connected to the internet.
- 2 Select **Annotations > Update Technology Annotations > From Agilent eArray** from the main menu.

The **Update eArray Technology** dialog box opens.

- 3 Provide your Agilent eArray **User Name** and **Password**.
- 4 Select the technology you want to update from the **Technology** drop-down list. This drop-down list contains all the standard Agilent technologies you have previously created.
- 5 Click **OK**.

GeneSpring accesses the Agilent eArray portal and displays the available annotations for the selected technology in the **Update Annotations of Technology (Step 3 of 3)** dialog box (see [Figure 2](#) on page 20).



**Figure 2** Update Annotations of Technology (Step 3 of 3) dialog box - available annotations for Agilent.SingleColor.12097

- 6 Select a suitable column from the existing technology for matching entities in the annotation source from the **Choose technology column to match** drop-down list.
- 7 Select the matching column in the annotation source from the **Choose file/genome column to match** drop-down list. For example,

the ID column in the GeneSpring technology matches the ProbeID column in an Agilent microarray data library.

**8** Determine whether you want to

- overwrite existing annotation information with the annotations retrieved from the eArray portal,
- append to the existing information, or
- fill in the missing values

from the **Choose Update Method** drop-down list.

**9** If the annotations listed in this window include Entrez Gene ID annotations, you have to change the Attribute Type and Data Type in that order as follows:

- a** Left-click on the word **Continuous** in the **Attribute Type** column. From the drop-down list that appears, select **Categorical**.
- b** Left-click on the word **integer** in the Data Type column. From the drop-down list that appears, select **string**.

**10** By default, only annotations for which GeneSpring could identify a matching **Column Mark** are selected. If you want to include other listed annotations it is recommended to assign a column mark before proceeding with the update.

To assign a column mark, select the row corresponding to the annotation you want to include. Left-click on **None** in the **Column Mark** column and select the matching mark from the drop-down list. If the desired term is not available from the drop-down list, select **Enter New** and type the new term.

**11** Click **Finish**.

GeneSpring updates the technology according to the choice you made in Step 8. A **Technology Inspector** opens.

**12** Review the details provided in the inspector.

**13** Click **Configure Columns** to add or remove columns from the **Entities** spreadsheet displayed in the inspector in the **Select Annotation Columns** dialog box. Click **OK**.

**14** Click **OK** to close the Technology Inspector.

The technology update from the Agilent eArray portal is complete.

### To update technology annotations from a GeneSpring update file

If you were not able to update technology annotations for one of the standard technologies available from the Agilent server, you can request a GeneSpring update file from [GeneSpring technical support](#). Once you have saved the update file on your computer, follow these steps to perform the update:

- 1 Select **Annotations > Update Technology Annotations > From GeneSpring Update File** from the main menu.

The **Please Select A .update File** dialog box opens.

- 2 The remaining steps are the same as those for creating a technology from a GeneSpring update file. See [“To create a technology from GeneSpring Update File”](#) on page 11 for the remaining steps.

## To update technology annotations from file

GeneSpring allows you to add annotations to a custom or standard technology beyond the annotations that were included when the technology was first created. For GeneSpring to be able to import these annotations, they have to present in a tab-delimited file format (\*.csv, \*.tsv, \*.xls, \*.txt).

- 1 Select **Annotations > Update Technology Annotations > From file or Biological Genome** from the main menu.

The **Update Annotations of Technology** wizard-driven workflow opens.

- 2 In Step 1 of 3 of this workflow, select the technology you want to update, indicate that you want to update from file and point to the file that contains the annotations.

- a Select the technology you want to update from the **Choose technology** drop-down list. This drop-down list contains all the technologies that were previously created.
- b Under **Choose source**, select **Update from file**.
- c Click **Browse** to locate the tab-delimited file that contains the annotations you want to add to the selected technology. Click **Open**.
- d Click **Next >>**.

- 3 In Step 2 of 3 of this workflow, provide details about the file format.

The **Preview** table indicates how GeneSpring read the selected input file by default. If the data columns were not recognized correctly, change the settings for the **Separator**, **Text qualifier**, **Missing value indicator**, and **Comment indicator** to match the specifications of the input file.

Click **Next >>**.

- 4 In Step 3 of 3 of this workflow, provide the necessary information for GeneSpring to correctly match the new annotations with the existing annotations in the selected technology.
  - a Select a suitable column from the existing technology for matching entities in the annotation source from the **Choose technology column to match** drop-down list.

- b** Select the matching column in the annotation source from the **Choose file/genome column to match** drop-down list.
- c** Determine whether you want to
  - overwrite existing annotation information with the annotations retrieved from the eArray portal,
  - append to the existing information, or
  - fill in the missing values

from the **Choose Update Method** drop-down list.

- d** If the annotations listed in this window include Entrez Gene ID annotations, you have to change the Attribute Type and Data Type in that order as follows:
  - Left-click on the word **Continuous** in the **Attribute Type** column. From the drop-down list that appears, select **Categorical**.
  - Left-click on the word **integer** in the Data Type column. From the drop-down list that appears, select **string**.

Click **Finish**.

- e** By default, only annotations for which GeneSpring could identify a matching **Column Mark** are selected. If you want to include other listed annotations it is recommended to assign a column mark before proceeding with the update.

To assign a column mark, select the row corresponding to the annotation you want to include. Left-click on **None** in the **Column Mark** column and select the matching mark from the drop-down list. If the desired term is not available from the drop-down list, select **Enter New** and type the new term.

GeneSpring updates the technology according to the choice you made in Step 4c. A **Technology Inspector** opens.

- 5** Review the details provided in the inspector.
- 6** Click **Configure Columns** to add or remove columns from the **Entities** spreadsheet displayed in the inspector in the **Select Annotation Columns** dialog box. Click **OK**.
- 7** Click **OK** to close the Technology Inspector.

The technology update from file is complete.



## To update technology annotations from biological genome

A biological genome is a set of known annotations for a particular organism. The NCBI provides biological genomes for a number of organisms, which can be imported into GeneSpring and used to augment at a standard or custom technology.

- 1 Create the biological genome for your research organism as described in [“How to create a biological genome in GeneSpring”](#) on page 27.
- 2 Select **Annotations > Update Technology Annotations > From file or Biological Genome** from the main menu.

The **Update Annotations of Technology** wizard-driven workflow opens.

- 3 In Step 1 of 3 of this workflow, select the technology you want to update and indicate that you want to update from Biological Genome.
  - a Select the technology you want to update from the **Choose technology** drop-down list. This drop-down list contains all the technologies that were previously created.
  - b Under **Choose source**, select **Update from Biological Genome**.
  - c Click **Next >>**.

GeneSpring retrieves the annotations from the previously created Biological Genome and presents the annotations in Step 3 of 3 of the workflow.

Step 2 is not displayed in this process, as the annotations were already stored in such a way that GeneSpring can import them correctly when the Biological Genome was created.

- 4 In Step 3 of 3 of the workflow, provide the necessary information for GeneSpring to correctly match the new annotations with the existing annotations in the selected technology.
  - a Select a suitable column from the existing technology for matching entities in the annotation source from the **Choose technology column to match** drop-down list.
  - b Select the matching column in the annotation source from the **Choose file/genome column to match** drop-down list.

- c Determine whether you want to
  - overwrite existing annotation information with the annotations retrieved from the eArray portal,
  - append to the existing information, or
  - fill in the missing values

from the **Choose Update Method** drop-down list.

- d By default, only annotations for which GeneSpring could identify a matching **Column Mark** are selected. If you want to include other listed annotations it is recommended to assign a column mark before proceeding with the update.

To assign a column mark, select the row corresponding to the annotation you want to include. Left-click on **None** in the **Column Mark** column and select the matching mark from the drop-down list. If the desired term is not available from the drop-down list, select **Enter New** and type the new term.

- e Click **Finish**.

GeneSpring updates the technology according to the choice you made in Step 4c. A **Technology Inspector** opens.

- 5 Review the details provided in the inspector.
- 6 Click **Configure Columns** to add or remove columns from the **Entities** spreadsheet displayed in the inspector in the **Select Annotation Columns** dialog box. Click **OK**.
- 7 Click **OK** to close the Technology Inspector.

The technology update from biological genome is complete.

## How to create a biological genome in GeneSpring

The NCBI provides a number of biological genomes for several organisms, which can be downloaded from the NCBI website. GeneSpring allows you to import these biological genomes in such a way, that the annotations it contains can be used to augment standard and custom technologies that you previously created in GeneSpring (see [“To update technology annotations from biological genome”](#) on page 25).

Since most standard technologies include the majority of the information contained in a biological genome, this feature is particularly useful for custom technologies with few annotations.

Biological genomes are independent of chip types and technologies. Therefore, once a biological genome has been created it can be used to augment technologies regardless of the chip type or vendor.

There are two ways to create a biological genome in GeneSpring:

- download directly from the NCBI ftp site.
- from a local folder on your computer.

The exact steps for both options are described in the following sections.

### To create a biological genome via direct download from the NCBI ftp site

GeneSpring is able to download biological genomes directly from the NCBI ftp site and create the genomes for the selected organisms in one step. In this process, GeneSpring first downloads the data for all organisms and then creates the genomes for the selected organisms. This typically takes a long time, depending on your internet connection, due to the large amount of data (around 300Mb) that has to be downloaded and processed.

#### NOTE

You cannot use GeneSpring while the biological genome is downloading from the NCBI ftp site using this option. If you want to use GeneSpring and download the biological genome files at the same time, follow the instructions [“To create a biological genome from a local folder”](#) on page 30.

- 1 Select **Annotations > Create Biological Genome** from the main menu.

The **Create Biological Genome** dialog box opens.

- 2 Under **Biological Genome**, select the organism, or organisms for which you want to create a biological genome.
- 3 Under **Genomic Data**, select **Download from NCBI ftp site**.
- 4 Click **OK**.

A Confirmation message opens.

- 5 Read the message carefully. If you want to go ahead with the direct download, click **Yes**.

The **Choose a folder** dialog box opens.

- 6 Click **Browse** to specify the folder in which you want to save the files that constitute the biological genome. Click **Open**.

#### NOTE

Choose the location and name of this folder carefully. If you want to create biological genomes for other organisms at a later time, you will have to specify this folder in the process.

- 7 Click **OK**.

GeneSpring downloads the biological genome files from the NCBI ftp site, saves them in the specified location and creates the biological genomes for the selected organisms.

All the information required to create the biological genome of any organism listed in the **Create Biological Genome** window is now saved on your computer. If you want to create biological genomes for other organisms at a later time, use the option [“To create a biological genome from a local folder”](#) on page 30.

### To create a biological genome from a local folder

To create biological genomes from a local folder you have to download the corresponding files from the NCBI ftp site and save them on your computer first. These files contain all the information required to create the biological genomes for any of the organisms listed in the **Create Biological Genome** dialog box.

Choosing this option for creating a biological genome, allows you to continue to work in GeneSpring while the files are downloading to your computer.

1 GeneSpring uses the following files from the NCBI ftp site:

- All\_Data.gene\_info,
- gene2accession,
- gene2go,
- gene2refseq,
- gene2unigene.

Download these files from the corresponding ftp addresses, which you can find under **Tools > Options > Miscellaneous > NCBI ftp URL**.

2 Save all these files in the same folder on your computer.

#### NOTE

Choose the location and name of this folder carefully. If you want to create biological genomes for other organisms at a later time, you will have to specify this folder in the process.

3 Select **Annotations > Create Biological Genome** from the main menu, when you have downloaded all the files.

The **Create Biological Genome** dialog box opens.

4 Under **Biological Genome**, select the organism, or organisms for which you want to create a biological genome.

5 Under **Genomic Data**, select **Use from local folder**.

6 Click **OK**.

The **Choose a folder** dialog box opens.

**7** Click **Browse** to specify the folder in which you saved the files that constitute the biological genome. Click **Open**.

**8** Click **OK**.

GeneSpring creates the biological genomes for the selected organisms. Repeat **Steps 3 to 8** to create biological genomes for other organisms.

## Downloading and Updating Annotations

Agilent provides genome annotations for many organisms through the Annotations Manager. These annotations are obtained from various sources like UCSC (<http://hgdownload.cse.ucsc.edu/downloads.html>), NCBI (<ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/>) and ENSEMBL (<http://asia.ensembl.org/index.html>).

Some of these annotations are required for experiment creation and have to be downloaded before an experiment is created. For example, reference, genes and transcripts, and target annotations are required for NGS SureSelect experiment creation.

Other annotations are required to perform different types of analysis in GeneSpring. For example, HomoloGene Groups are required to perform Multi-Omic Pathway Analysis to map entities across different organisms. Another example is the dbSNP annotation, which allows you to distinguish between known and novel SNPs following SNP Detection Analysis.

The following sections describe in detail the required steps for downloading these annotations and updating them when updates become available.

### How to download annotations in the Annotations Manager

Using the Annotations Manager, there are two ways to bring sequence-based annotations into GeneSpring:

- Download from the Agilent server.
- Import from a GeneSpring update file.

The steps for each of these options are described in the following sections.



### To download annotations from the Agilent server

Downloading annotations from the Agilent server requires an internet connection. Due to the large size of some of the organism's annotations, this process can take some time depending on the speed and download limitations of your internet connection.

- 1 Make sure you are connected to the internet.
- 2 Select **Annotations > Annotations Manager** from the main menu.

The **Annotations Manager** window opens.

#### NOTE

If you have not previously downloaded any of the annotations available in this window, the complete list of organisms and their annotations is displayed in the left pane. Once you have downloaded annotations, the Annotations Manager only displays these annotations the next time you open it. To display all the other available organisms and their annotations click **List**.

- 3 Click the name of any organism or annotation type to view its description in the pane on the right.
- 4 You can either download all or selected annotations for one or several organisms at the same time:
  - Select the organism name to download all available annotations for that organism.
  - Expand the hierarchy for the desired organism and builds and select only those annotations that you want to download.

The total download size and required disk space for the selected annotations are displayed in the bar below the left and right panes, respectively.

- 5 Click **Update**.

GeneSpring downloads the selected annotations from the Agilent server. Only the downloaded organisms and annotations appear in the left pane of the window when the download is complete.

- 6 Click **Close**.

The imported annotations are now available for creating experiments, display in the Genome Browser, or performing analyses in GeneSpring.

## Downloading and Updating Annotations

If you are not able to download annotations from the Agilent server, you can import them from a GeneSpring update file (see [“To import annotations from a GeneSpring update file”](#) on page 35).

### To import annotations from a GeneSpring update file

If you are not able to download annotations from the Agilent server (see “To download annotations from the Agilent server” on page 33), contact [GeneSpring technical support](#) and ask for the corresponding \*.update or \*.zip update files. You will either receive a download link to save the files on your computer, or as an email attachment, depending on the size of the annotations. Once you have saved these files on your computer, you do not require an internet connection to import the annotations.

- 1 Select **Annotations > Annotations Manager** from the main menu.

The **Annotations Manager** window opens. Only annotations that you previously downloaded are listed in the pane on the left.

- 2 Click the drop-down arrow next to **List** and select **From File**.

The **Open** dialog opens.

- 3 Browse for the \*.update or \*.zip file you saved on your computer earlier. Click **Open**.

The annotations contained in the update file appear in the left pane of the Annotations Manager.

- 4 You can either download all or selected annotations for one or several organisms at the same time:

- Select the organism name to download all available annotations for that organism.
- Expand the hierarchy for the desired organism and builds and select only those annotations that you want to download.

The total download size and required disk space for the selected annotations are displayed in the bar below the left and right panes, respectively.

- 5 Click **Update**.

GeneSpring imports the selected annotations from the update file. When the download is complete, only the imported organisms and annotations appear alongside any previously downloaded or imported organisms in the left pane of the window.

- 6 Click **Close**.

The imported annotations are now available for creating experiments, display in the Genome Browser, or performing analyses in GeneSpring.

## How to update annotations in the Annotations Manager

As research progresses, annotations are updated by the curators at NCBI, UCSC, ENSEMBL and other institutes that provide annotations. Whenever such updates become available on the source websites, Agilent applies these updates on its server and makes them available to download through the Annotations Manager.

When the Annotations Manager is launched for the first time, all organism names, builds, and annotations are marked with an asterisk to indicate that the item contains information that is not yet present in GeneSpring. After downloading the annotations to GeneSpring, the asterisk disappears. GeneSpring notifies you when an update becomes available for one of these previously downloaded annotations through an Update message that appears during start up.

To update these annotations follow the same steps as described in [“To download annotations from the Agilent server”](#) on page 33. After you click **List**, the organism name, build, and annotations for which updates became available now appear once again marked with an asterisk.

If you are not able to update the annotations from the Agilent server, ask [GeneSpring technical support](#) for the corresponding update files and follow the same steps as described in [“To import annotations from a GeneSpring update file”](#) on page 35.

# Downloading and Importing Gene Sets

This section describes the required steps for downloading gene sets from the Broad Institute and Stanford University websites before importing them into GeneSpring.

## How to download Broad GSEA gene sets

Downloading Broad Institute gene sets for gene set enrichment analysis (GSEA) is free, but you have to register with the website to gain access.

- 1 Go to <http://www.broadinstitute.org/gsea/msigdb/index.jsp>.
- 2 Under **Overview**, click **Download** gene sets.
- 3 Click the file name of the gene set that you want to download, or download all gene sets at once.

GeneSpring supports the import of gene sets in one of the following file formats:

- \*.txt/\*.tsv - The first row contains headers, the remaining rows contain one gene each.
- \*.grp - Each row contains a gene.
- \*.gmt - Gene Matrix Transposed file format in which each row contains a gene set.
- \*.xml - Molecular signature database file format (msigdb\_\*.xml).

### NOTE

To perform GSEA the gene set files you download have to contain the identifiers for the annotations that are present in the technology of your experiment. For Affymetrix, Agilent, and Illumina standard technologies, Agilent provides both Gene Symbol and Entrez Gene ID annotations. For custom technologies, these annotations have to be provided when creating the technology (see [“How to create a technology”](#) on page 7). We recommend to download the gene symbol files for best results.

- 4 Save the file on your computer.

## Downloading and Importing Gene Sets

- 5 Repeat Steps 3 and 4 until you have downloaded all the gene sets that you are interested in.

Once you have downloaded all the gene sets you want to use for GSEA, you can import them into GeneSpring as described in [“How to import gene sets into GeneSpring”](#) on page 40.

## How to download gene sets from the Stanford University website

Downloading gene sets from the Stanford University website is free.

- 1 Go to <http://www-stat.stanford.edu/~tibs/GSA/>.
- 2 On the page, scroll down to **Available gene set collections**.

The first item in the list is a link to the gene sets provided by the Broad Institute. See “[How to download Broad GSEA gene sets](#)” on page 37 for instructions on how to download these gene sets.

The remaining items are gene sets provided by Stanford University.

- 3 Right-click the name of the gene set that you want to download and select **Save Link As**.

Depending on your browser, a dialog box opens that allows you to choose the location for saving the file and providing a name (other than the default name) for the gene set you are about to download.

- 4 Click **Save**.
- 5 Repeat Steps 3 and 4 until you have downloaded all the gene sets you want to work with.

Once you have downloaded all the gene sets you want to use for GSEA, you can import them into GeneSpring as described in “[How to import gene sets into GeneSpring](#)” on page 40.

## How to import gene sets into GeneSpring

Once you have downloaded all the gene sets you want to use for GSA or GSEA, follow these steps to import them into GeneSpring.

- 1 In an open experiment, select **Tools > Import BROAD GSEA Gene sets** from the main menu.

The **Choose file for importing** dialog box opens.

- 2 Select the appropriate file type from the **Files of type** drop-down list.
- 3 Browse for the gene set file that you want to import. Click **Open**.

The **More information required** dialog box opens.

- 4 Under **Choose MIT-BROAD Gene Set Group**, specify the group of gene sets from the Broad Institute you are importing.

If you are importing a gene set from any source other than the Broad Institute, select **Custom List** and provide additional details in the **Additional Notes** field if desired.

- 5 Click **OK**.

GeneSpring imports the gene sets. When the import is complete, an Information dialog tells you how many gene sets were saved successfully.

- 6 Click **OK**.

- 7 Repeat Steps 1 to 7 until you have imported all the gene sets that you wanted to work with in GeneSpring.

You are now ready to perform GSA or GSEA, available from the **Results Interpretation** section of the **Workflow Browser** in GeneSpring.



## Downloading and Updating Interaction Databases

This section describes in detail the steps required to download interaction databases for the first time and update them when updates become available.

### How to download GeneSpring interaction databases

GeneSpring interaction databases are available to download from the Agilent server. If you are not able to download the interaction database for your organism from the server, you can import the database from a GeneSpring update file. The required steps for both options are described in the following sections.

Interaction databases are saved in the GeneSpring installation folder by default. Before you start downloading interaction databases, please note that they can require a large amount of disk space depending on the organism. For example, the human interaction database requires around 6GB of disk space.

### To download interaction databases from the Agilent server

- 1 Make sure that you are connected to the internet.
- 2 Select **Annotations > Update Pathway Interactions > From Agilent Server** from the main menu.

If you are attempting to download interaction databases for the first time, GeneSpring prompts you to download the Pathway Infrastructure Database. This database is required for downloading any organism-specific interaction database.

- 3 Click **Yes** to continue.

The **Automatic Software Update** window appears when the Pathway Infrastructure Database download is complete. It lists all available interaction databases.

- 4 Select the interaction databases that you want to download.

Details for each selected interaction database appear in the lower panel of the window.

- 5 Click **Update**.

An **Automatic Software Update** information dialog informs you of the amount of hard disk space that is required to download the selected interaction databases.

- 6 Click **OK**.

GeneSpring downloads the interaction databases from the Agilent server. An **Automatic Software Update** information dialog confirms when the download is complete.

- 7 Click **OK**.

The selected interaction databases are now available for all types of Pathway Analysis in GeneSpring.

To inspect the number of entities by entity type and the number of relations by relation type available for each organism and in the generic database (organism-independent metabolites) select **Annotations > Pathway Database Statistics** from the main menu.

Once an organism's interaction database has been downloaded it will no longer appear listed in the Automatic Software Update window until an update becomes available.

### To import interaction databases from a GeneSpring update file

If you were not able to download interaction databases from the Agilent server (see [“To download interaction databases from the Agilent server”](#) on page 42) contact [GeneSpring technical support](#). They can provide you with a GeneSpring \*.update or \*.zip update file for the same list of organisms available from the Agilent server. Save the update file to your computer.

If you are importing interaction databases for the first time, you have to first import the update file for the Pathway Infrastructure Database before importing the update file for any organism-specific interaction databases.

- 1 Select **Annotations > Update Pathway Interactions > From Update File...** from the main menu.

The **Please Select A .update File** dialog box opens.

- 2 Locate the \*.update file or \*.zip archive on your computer.
- 3 Click **Open**.

The interaction database contained in the update file appears in the **Automatic Software Update** window.

- 4 Select the interaction database.

A summary of the selected interaction database appears in the bottom pane of the window.

- 5 Click **Update**.

An **Automatic Software Update** information dialog informs you of the amount of hard disk space that is required to import the selected interaction database.

- 6 Click **OK**.

GeneSpring extracts the interaction database from the update file. An **Automatic Software Update** information dialog confirms when this process is complete.

- 7 Click **OK**.

The selected interaction databases are now available for all types of Pathway Analysis in GeneSpring.

## Downloading and Updating Interaction Databases

To inspect the number of entities by entity type and the number of relations by relation type available for each organism and in the generic database (organism-independent metabolites) select **Annotations > Pathway Database Statistics** from the main menu.

Once an organism's interaction database has been downloaded it will no longer appear listed in the Automatic Software Update window until an update becomes available.

## How to update GeneSpring interaction databases

The GeneSpring Interaction Database is periodically updated with relations derived from the latest PubMed abstracts. When updates are available, GeneSpring notifies you with a message the next time you launch the application.

You can update GeneSpring interaction databases from the Agilent server or a GeneSpring update file. The steps for each of these processes are the same as those described in [“To download interaction databases from the Agilent server”](#) on page 42 and [“To import interaction databases from a GeneSpring update file”](#) on page 43. The only difference is that you do not have to download the Pathway Infrastructure Database.

### To update pathways following an interaction database update

Every time you update an organism’s interaction database, GeneSpring presents all the pathways that are affected by the update in the **Update Pathways** dialog. You then have the option to update or delete these pathways.

- 1 Select each pathway in turn to review which entities were updated and which entities became obsolete. The selected pathway and entities are displayed in the pathway viewer on the right.
  - Click **Display Obsolete** to view the entities that will be deleted from the pathway if you update the pathway.
  - Click **Display Updated** to view the entities that will be updated if you choose to update the pathway.
  - Click **Display All** to display all the entities in the current version of the pathway.
- 2 Select the pathways that you want to delete. Click **Delete**.
- 3 Select the pathways that you want to update. Click **Update**.
- 4 Click **Close** when you have completed your review of pathways affected by an interaction database update.

If you do not want to update or delete any of the listed pathways at this time you can return to the **Update Pathways** dialog at a later time by selecting **Annotations > Update Pathways** from the main menu.

## Downloading and Updating BridgeDb and ID Mapper Databases

This section describes in detail the steps required to download BridgeDb and ID Mapper databases for the first time and update them when updates become available.

### How to download BridgeDb and ID Mapper databases

BridgeDb and ID Mapper databases are available to download from the Agilent server for a number of organisms. The human BridgeDB and the metabolite ID Mapper databases are already packaged with your GeneSpring installation and do not have to be downloaded again.

If you are not able to download a BridgeDb or ID Mapper database for your organism from the server, you can import the database from a GeneSpring update file available from GeneSpring Support. The required steps for both options are described in this section.

### To download BridgeDb or ID Mapper databases from the Agilent server

- 1 Make sure that you are connected to the internet.
- 2 Select **Annotations > Update BridgeDb > From Agilent Server** from the main menu.
- 3 Click **Yes** to continue.

The **Automatic Software Update** window opens. It lists all available BridgeDb and ID Mapper databases.

- 4 Select the databases that you want to download.

Details for each selected database appear in the lower panel of the window.

- 5 Click **Update**.

An **Automatic Software Update** information dialog informs you of the amount of hard disk space that is required to download the selected databases.

- 6 Click **OK**.

GeneSpring downloads the databases from the Agilent server. An **Automatic Software Update** information dialog confirms when the download is complete.

- 7 Click **OK**.

The selected databases are now available for all types of Pathway Analysis in GeneSpring.

Once an organism's BridgeDb or ID Mapper database has been downloaded it will no longer appear listed in the Automatic Software Update window until an update becomes available.

### To import BridgeDb or ID Mapper databases from a GeneSpring update file

If you were not able to download BridgeDb or ID Mapper databases from the Agilent server (see [“To download BridgeDb or ID Mapper databases from the Agilent server”](#) on page 47) contact [GeneSpring technical support](#). They can provide you with a GeneSpring \*.update or \*.zip update file for the same list of organisms available from the Agilent server. Save the update file to your computer.

- 1 Select **Annotations > Update BridgeDb > From Update File...** from the main menu.

The **Please Select A .update File** dialog box opens.

- 2 Locate the \*.update file or \*.zip archive on your computer.
- 3 Click **Open**.

The database contained in the update file appears in the **Automatic Software Update** window.

- 4 Select the database.

Details of the selected database appear in the bottom pane of the window.

- 5 Click **Update**.

An **Automatic Software Update** information dialog informs you of the amount of hard disk space that is required to import the selected database.

- 6 Click **OK**.

GeneSpring extracts the database from the update file. An **Automatic Software Update** information dialog confirms when this process is complete.

- 7 Click **OK**.

The selected databases are now available for all types of Pathway Analysis in GeneSpring.

Once an organism's database has been downloaded it will no longer appear listed in the Automatic Software Update window until an update becomes available.



## How to update BridgeDb and ID Mapper databases

GeneSpring notifies you when updates to the BridgeDb and ID Mapper databases that you downloaded earlier become available on the Agilent server.

You can update these databases from the Agilent server or a GeneSpring update file. The steps for each of these processes are the same as those described in [“To download BridgeDb or ID Mapper databases from the Agilent server”](#) on page 47 and [“To import BridgeDb or ID Mapper databases from a GeneSpring update file”](#) on page 48.

## Getting help

### Online help

**Press F1** To get more information about a menu, toolbar, window, or dialog box, place the cursor on the part of the menu, toolbar, window, or dialog box of interest and press the **F1** key.

**Help menu** Click **Help > Documentation Index** to access the release notes, quick start guides, and HTML and PDF versions of the GeneSpring manual.

**Online support portal** Visit [www.genespring-support.com](http://www.genespring-support.com) to contact the GeneSpring support team and access additional demo data sets, user guides, and video tutorials.

**Webinars** Visit our [Webinar page](#) to register for one of our monthly webinars on available features in GeneSpring.

### Documents

**GeneSpring manual** Agilent GeneSpring [User Manual](#) (Agilent publication n/a, May 2013).

**www.agilent.com**

## **In this Book**

This book tells you how to import, download, and update the various types of data supporting the creation of experiments and analysis of sample data in Agilent GeneSpring.

© Agilent Technologies, Inc. 2013

Printed in United States of America  
05/2013



G9220-90000



**Agilent Technologies**